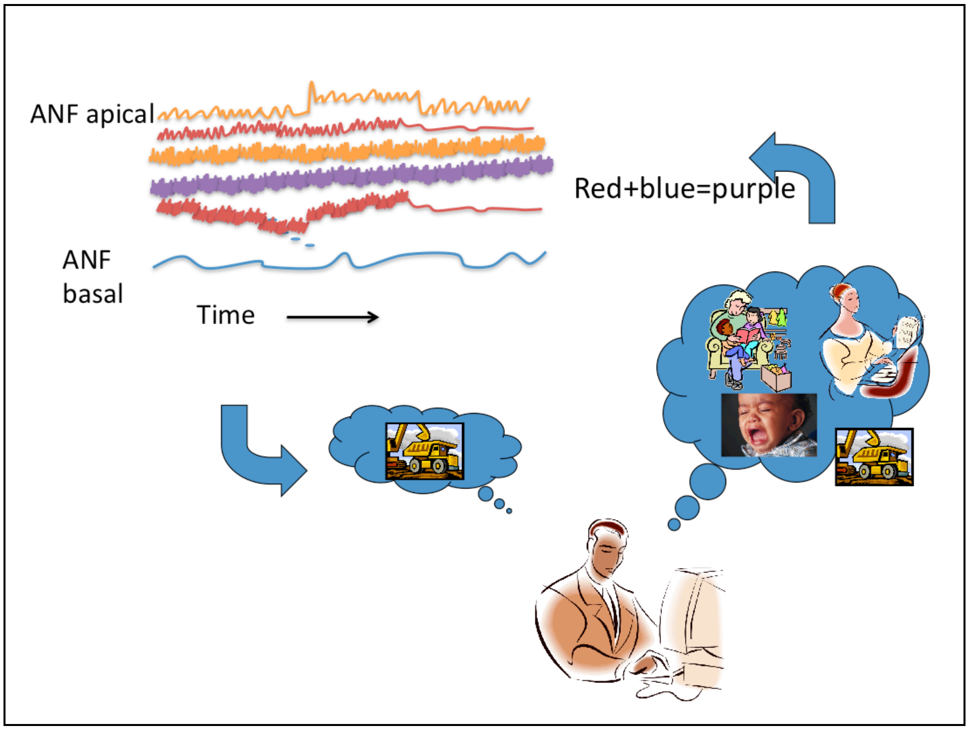
Sound source segregation (determination)

The process by which acoustic components are identified as coming from one or more sound sources.



We have already described this problem: There are multiple sound sources. Each one is making a complex sound. So the activity in some of these nerve fibers come from one source, some comes from another source, and some nerve fibers could be responding to multiple sources (the purple trace represents the response to components of the “red” source and the “blue: source mixed together. So the activity elicited by these multiple sources is interleaved across nerve fibers and frequently the same neurons are responding to multiple sources.

Nonetheless you can separate all this activity into separate sources– you can hear the baby crying and the truck and your wife reading to the kids and your mother playing the piano.. You can identify each of the sources and you can understand what your wife is reading. You know where the sounds are coming from. And you can listen to one of the sounds while ignoring the others, Unfortunately, the nerve fibers responses are not color coded so that the brain can figure out which activity is coming from which source, So how does the auditory nervous system take the response of the auditory nerve fibers and construct this auditory scene?

Two types of method for assessing sound source segregation

- Auditory steaming
- Thresholds

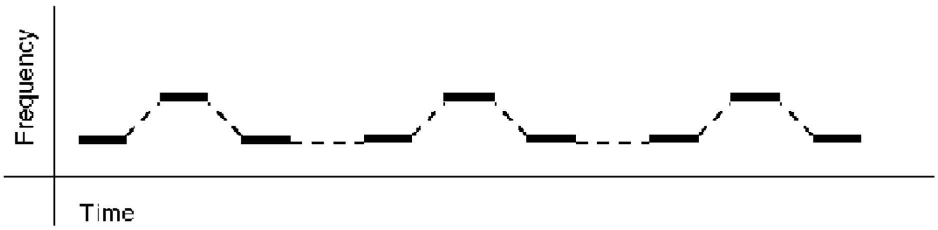
How do we know which cues people use?

Auditory streaming

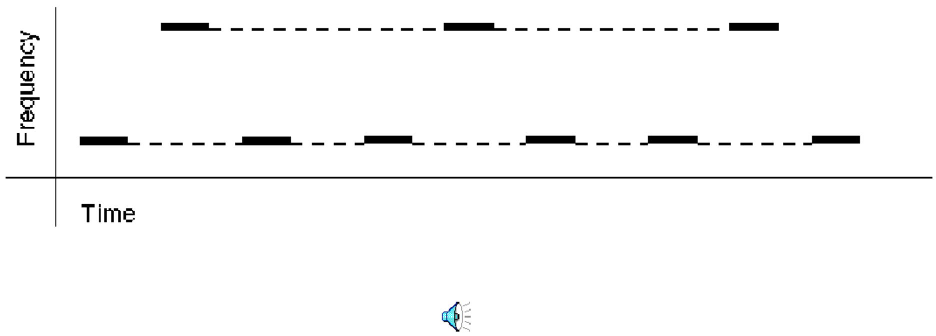
When two frequencies are close together, you hear them as one source, or “stream”; when they are far apart you hear two.

People tend to hear two streams as simultaneous rather than alternating when the streams are segregated. So you could ask people whether they hear two streams or one when I change the characteristics of the two sound sequences. People have trouble figuring out the temporal order of events in two streams. It sounds like the streams are occurring simultaneously.

A single sound source is perceived



Two sound sources are perceived



Disadvantages of auditory streaming method

- Frequency separation isn't the only thing that we use to separate sources; this is limited way to study the problem.
- "What does this sound like to you?" doesn't seem like the sort of question that would produce a reliable answer.

In fact, streaming methods have no way to control for response bias, for example.

Auditory streaming is a method for studying

- Sound source segregation
- Localization
- Lateralization
- Frequency discrimination

A disadvantage of the auditory streaming method is that

- only tones can be studied
- it uses a very subtle perception
- it does not produce very reliable results
- it cannot be used to study infants or children

If one sound can be segregated from another, then the threshold for the sound should be lower.

Acoustic cues that could be used to segregate components into sources

- Spectral separation
- Temporal separation
- Temporal onsets and offsets
- Spectral profile
- Harmonicity
- Spatial separation
- Temporal modulations

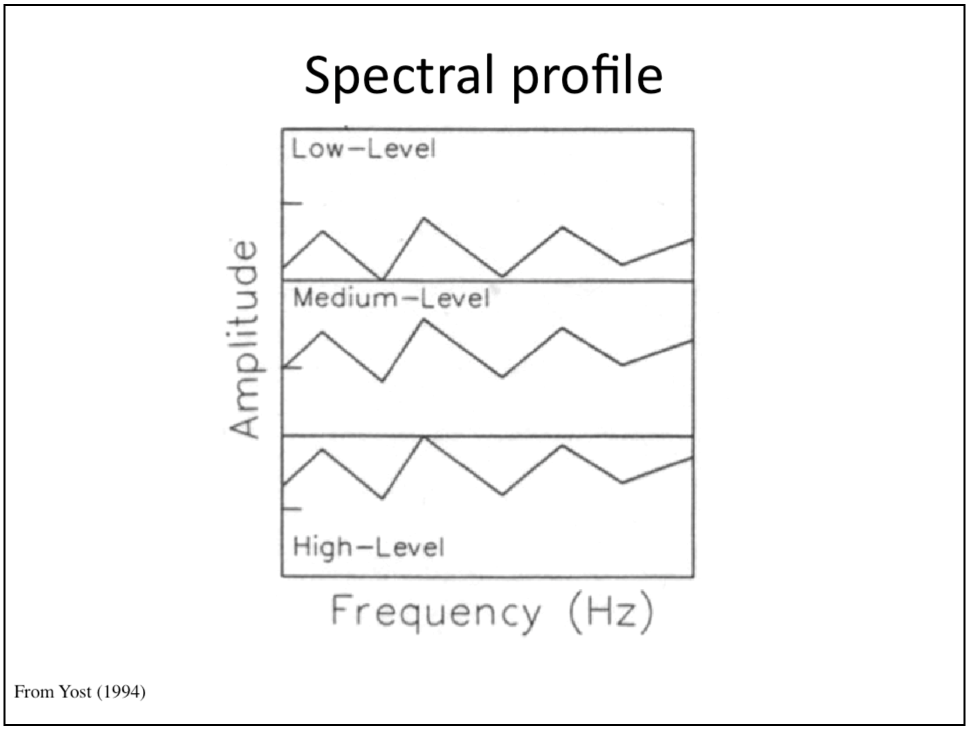
There are several sorts of acoustic information that the auditory system could use to figure out which components belong together. Most studies of sound source segregation focus on identifying and manipulating acoustic cues thought be responsible for sound source segregation. However, the auditory system rarely receives only one of these cues in isolation. Instead we have several cues available.

Some of these cues (the top three) are pretty self-explanatory, and we've already seen that when these cues are available to separate a probe from a masker, threshold is lower. For example, we know that when the probe and the masker have vey different frequencies– spectral separation, then the threshold for the probe is lower.

The same is true of temporal separation. We know that if the probe is close to the masker in time, we can get masking, but that as we increase the time difference between them, threshold for the probe gets lower.

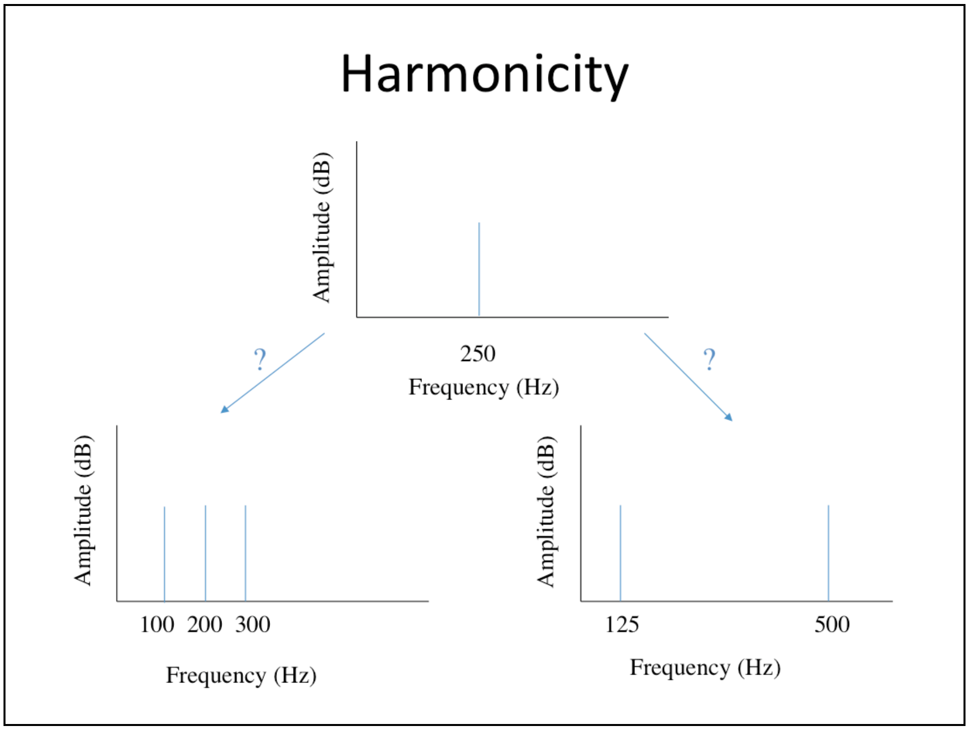
And we know that when sounds come on together– forward fringe masking== or go off together– backward fringe masking, threshold for the probe is higher. When they don't' come on or go off together, threshold goes down.

These other cues take a little more explanation,



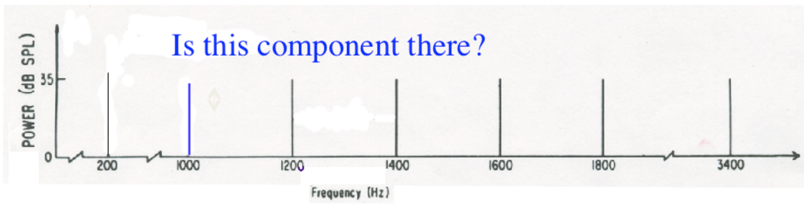
Spectral profile refers to the pattern of amplitudes across frequency--the spectral shape. The components of sound coming from a single source tend to maintain the same amplitude relationship to each other even when the overall amplitude changes. Thus, the spectral profile, or spectral shape, remains constant, and we could use that fact to group components together. The spectral profile is related to the timbre, or quality, of the sound.

People are actually very sensitive to the spectral profile of a sound. And we know that people are better at detecting a sound that has a different spectral profile than the masker.

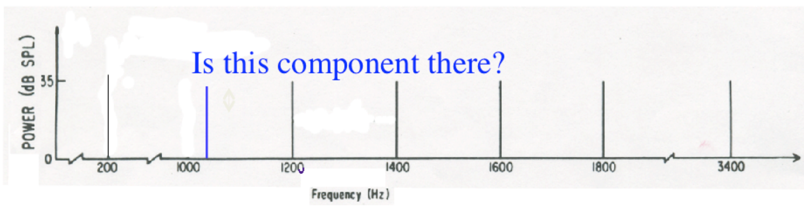


Sources that produce periodic sounds will produce harmonically related components. We could group a component with other components that have a common fundamental frequency. If you mistune an individual harmonic even just a little, listeners will segregate it from the ret of the complex.

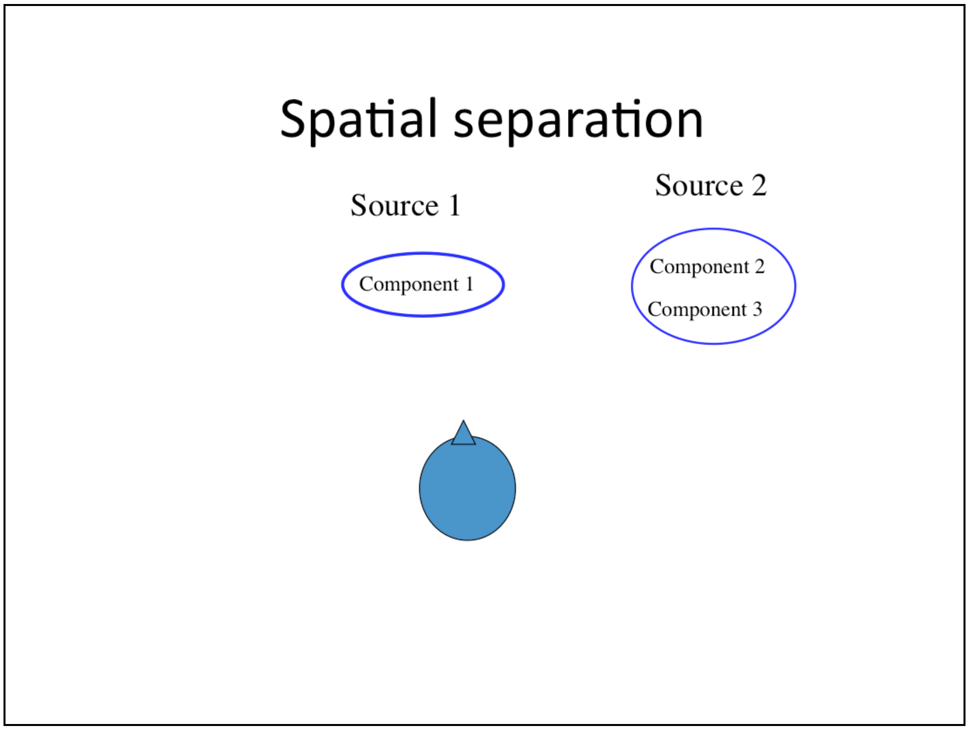
Harmonicity



Lower threshold

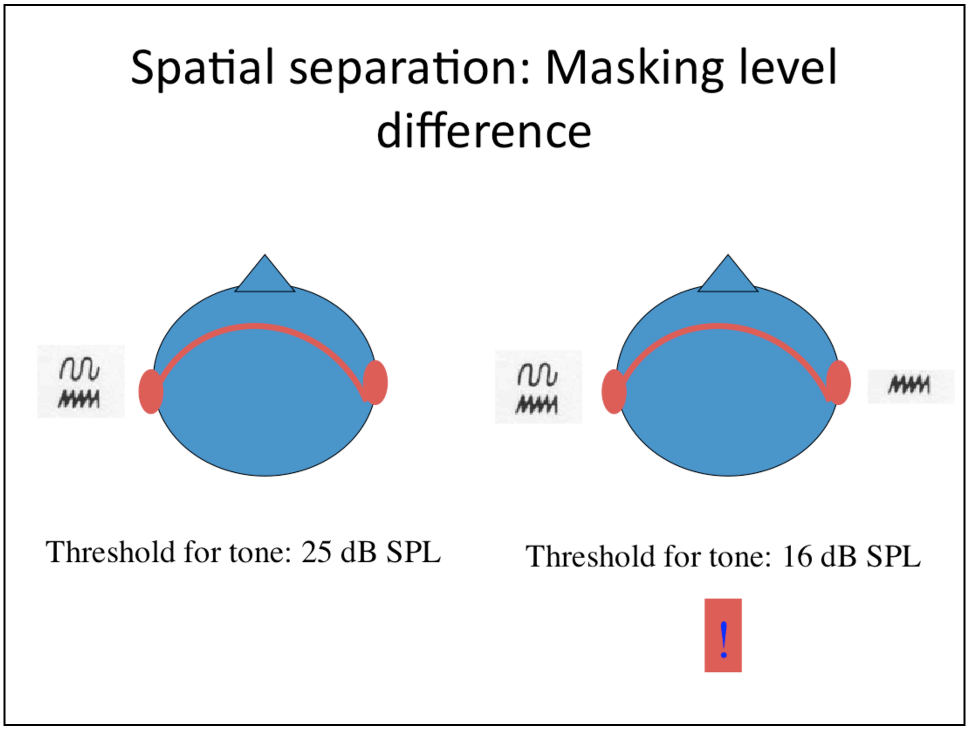


If we measure the detection threshold for a component of a harmonic complex, the threshold is lower if the component is mistuned.



We could group components together if they come from a common location in space.

Whether people use this cue in laboratory studies or not seems to depend on how many simultaneous sources are used. Spatial separation seems to become important when there are several (3 or more) sources.

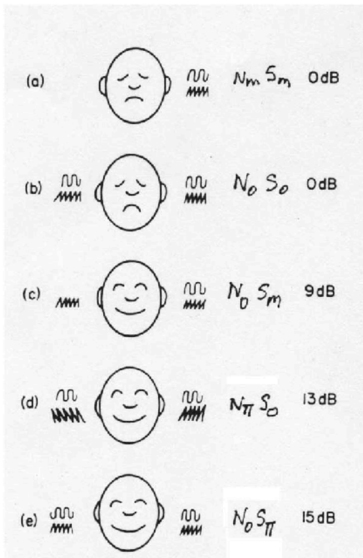


The effects of spatial separation have been studied in lateralization-type experiments. If we measure the threshold for a tone in noise when the tone and noise are presented to the same, single ear, the threshold is higher than if the noise is presented to both ears while the tone is presented to just one ear. This difference in threshold is referred to as the masking level difference.

Spatial separation: Masking level difference

If we think about the lateralization of the tone and noise, we can see why this might happen. When tone and noise are presented to one ear, they are lateralized to the same position in the head. If the noise is presented to both ears, it will be lateralized to the center of the head, while the tone, still in one ear, will be lateralized toward the ear of presentation.

Masking level difference



Modified from Gelfand (1998)

N= noise, S = signal

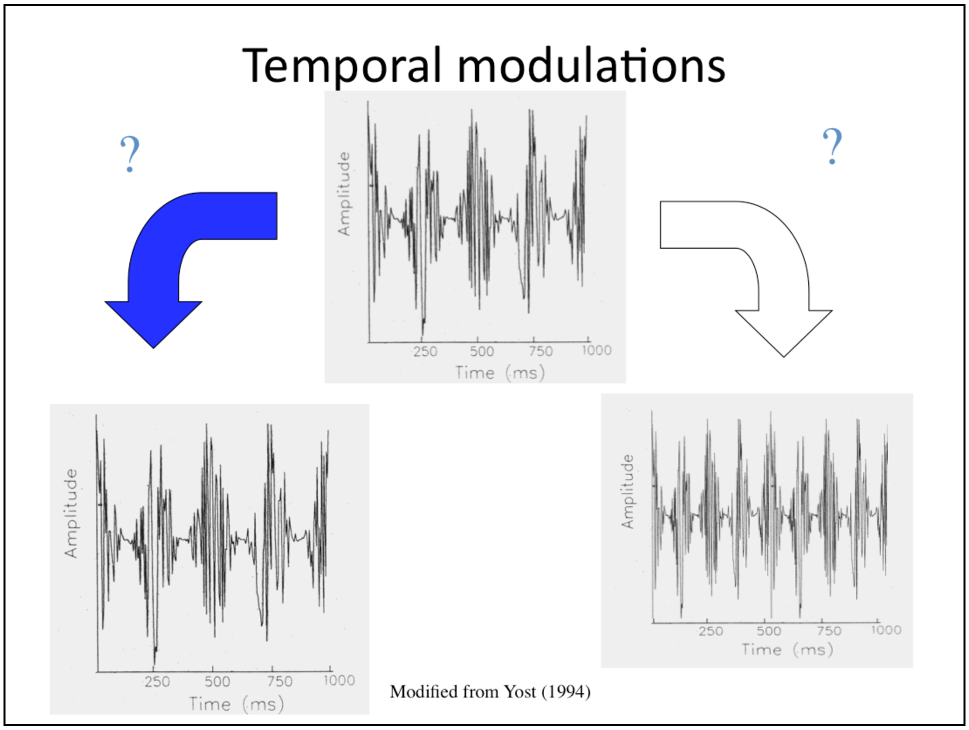
Monotic = one ear (m)

Diotic = 2 ears, same sound in both (0)

Dichotic = 2 ears, different sound in each (π)

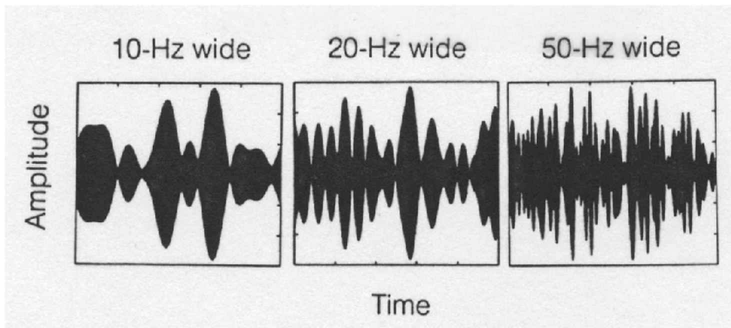
The MLD is the improvement in audibility that results from dichotic listening

The masking level difference occurs in various conditions and it varies in size with condition. In MLD terminology, N stands for noise and S stands for signal. A subscript "m" means that the sound is presented to one ear only (monotic). A subscript "0" (zero) means that the sound is presented to both ears and that it is identical in the two ears (diotic). A subscript "π" means that the sound is presented to both ears but that the sound in one ear is 180 degrees (π radians) out of phase with the sound in the other ear (dichotic). N_oS_Π produces the largest MLDs.



The components of sound from a common source change in amplitude together over time. So we would group a component with other components that have a common amplitude (temporal) modulation. Temporal modulation could also be frequency modulation, as when a player produces vibrato on a string instrument.

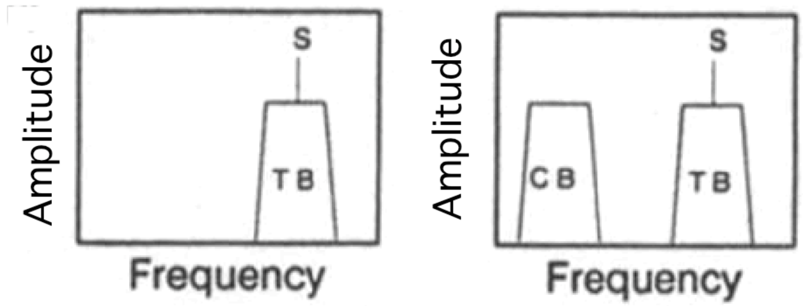
A noise band has a distinct amplitude envelope that we can isolate and apply to different carrier frequencies



From Yost (1994)

The average rate at which the amplitude varies is proportional to the bandwidth of the noise band. But more importantly, it is possible to extract the amplitude envelope of one noise band and impose it on another noise band. Now the two noise bands have the same amplitude envelope. We say they are comodulated. How should comodulated noise bands be perceived?

Temporal modulations: Comodulation Masking Release



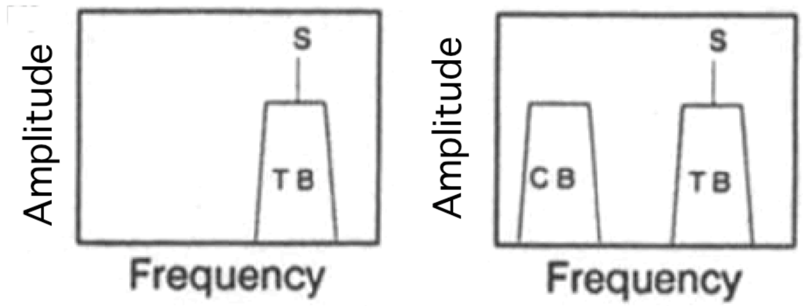
S = signal, a tone
TB = target band, a noise band
CB = cue band, another noise band

If CB is more than a critical band (auditory filter) away from TB in frequency, will the threshold for S change when CB is present?

From Yost (1994)

One experiment that demonstrates the role of temporal modulations in osund source segregation (as indicated by thresholds) is the comodulation masking release experiment. Threshold for a tone ("S" in the amplitude spectrum) us first measured when the tone is masked by a narrow band of noise centered on the tone frequency. The noise band centered on the tone frequency is called the target band ("TB" in the amplitude spectrum). Then the threshold for the tone is measured when both the target band and another noise band--the cue band--are present. If the cue band is processed by a different auditory filter than the target band, will its presence change the threshold for the tone?

Temporal modulations: Comodulation Masking Release



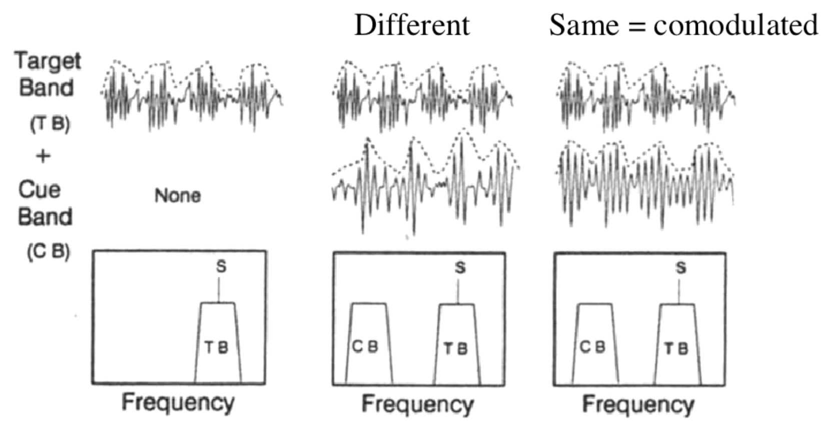
If CB is more than a critical band (auditory filter) away from TB in frequency, will the threshold for S change when CB is present?

IT DEPENDS.

From Yost (1994)

Sometimes the threshold for the tone is unaffected by the presence of the cue band. Sometimes the tone threshold is IMPROVED by the presence of the cue band.

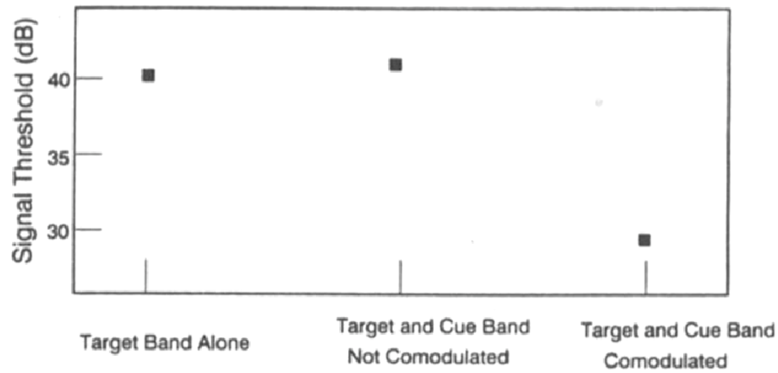
The target band and the cue band could have the same or different amplitude envelopes



From Yost (1994)

The envelopes are temporal modulations in amplitude. So noise bands with different envelopes should be perceived as coming from different sources, while noise bands with comodulated envelopes should be perceived as coming from the same source. The tone represents still another source, but perceptually grouping the target band with the cue band perceptually separates the target band from the tone and the tone should be easier to hear.

Comodulation of noise bands makes threshold for signal lower



Comodulation masking release is the improvement in audibility that results from common amplitude modulation across auditory filters.

From Yost (1994)

Adding a cue band without comodulation gives the same threshold for the tone as when only the target band is presented, as long as the cue band frequency is processed bu a different auditory filter as the target band. But if the cue and target bands are comodulated, threshold for the tone improves.

The assumption underlying the use of thresholds to study sound source segregation is

- thresholds are worse when masker and probe can be segregated
- thresholds are better when masker and probe can be segregated
- thresholds vary depending on how the listener listens.

The masking level differences demonstrates the importance of which of the following for sound source segregation?

- Temporal onsets
- Temporal modulation
- Spectral profile
- Spatial location

Which of these demonstrates the importance of temporal onsets to sound source segregation?

- Masking level difference
- Comodulation masking release
- Forward fringe masking
- Profile analysis

Comodulation masking release demonstrates the importance of which of the following for sound source segregation?

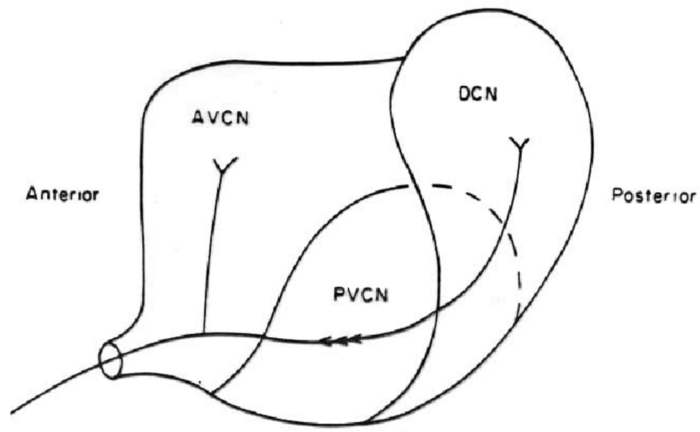
- Temporal onsets
- Temporal modulation
- Spectral profile
- Spatial location

How does the brain identify features of sound that allow us to segregate them?

- Spatial separation – calculation of IID, ITD in MSO and LSO
- Temporal separation
- Temporal onsets and offsets
- Temporal modulations
- Spectral profile
- Harmonicity
- Spectral separation

Well, spatial separation, we already know about. The rest of these cues, I've separated into two groups. The top group involves identifying the temporal characteristics of a sound, while the bottom group involves identifying the spectral characteristics– the amplitude spectrum of the sound.

The cochlear nucleus



From Pickles (1988)

Feature extraction begins at the first relay in the nervous system, the cochlear nucleus. The AVCN we've already discussed. We know that the neurons there are primarylike, like auditory nerve fibers. The other two divisions of the cochlear nucleus are the posteroventral cochlear nucleus or PVCN and the dorsal cochlear nucleus, or DCN. The PVCN is the first stage of temporal feature detection, while the DCN is first stage of spectral feature detection. Besides some special circuitry that we will not discuss, these two sections of the CN have neurons with very specialized structures that allow them to respond to very specific features of a sound.

Cell types in the cochlear nucleus

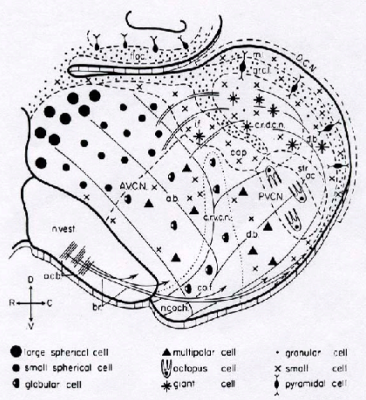
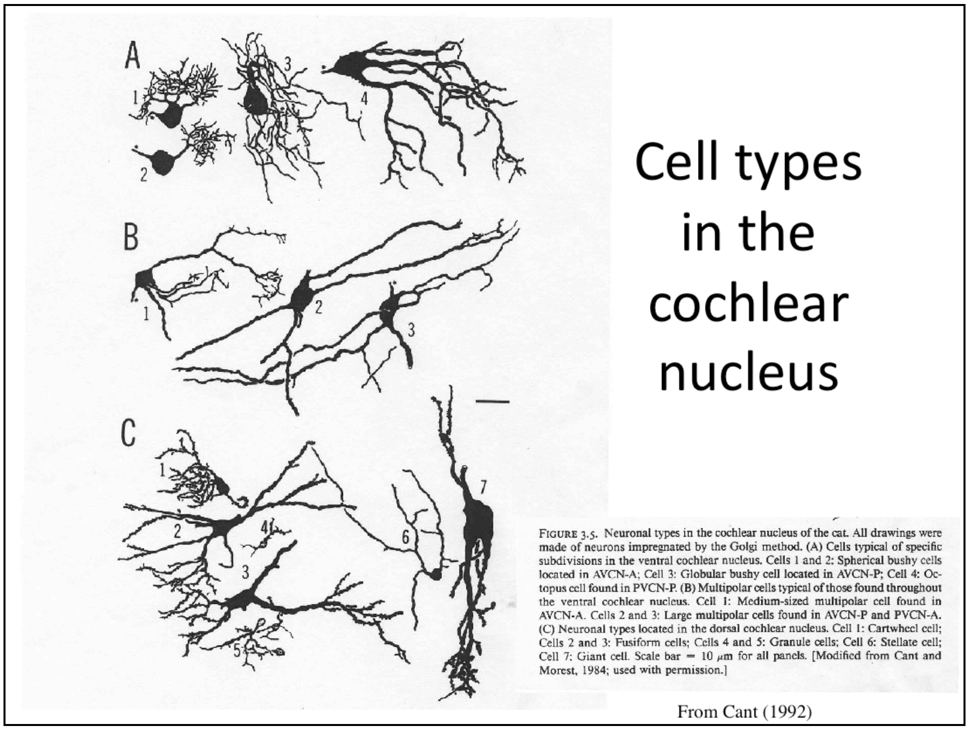


Fig. 6.2 A cytoarchitectural map of the cochlear nucleus is shown in sagittal section. The predominant cell types in each region are represented. AVCN, antero-ventral nucleus; cap, peripheral cap of small cells; crdcn, central region of DCN; crcvn, central region of ventral nucleus; DCN, dorsal nucleus; floc, flocculus (cerebellum); gcl, granular cell layer; if, intrinsic fibres; ml, molecular layer; PVCN, posteroventral nucleus; strac, dorsal and intermediate acoustic striae. From Osen and Roth (1969, Fig. 1).

From Pickles (1988)

Recall that numerous cell types are found in the cochlear nucleus with different neuron types being found in different subdivisions.



The shapes of these cells are not random. They are built the way they are to make them respond to only certain things about the message they receive from the auditory nerve.

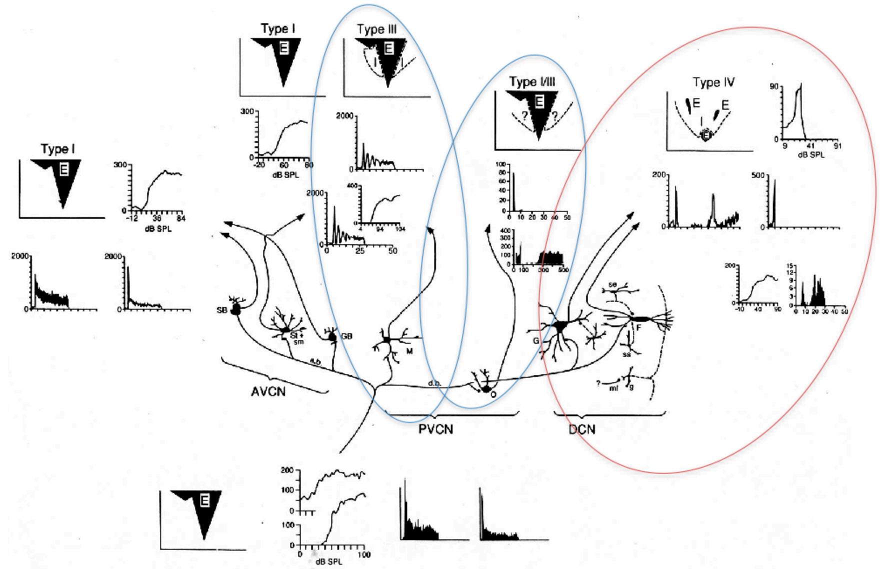


FIGURE 16.12 Cochlear nucleus cell types and the variety of physiological measures are shown. The cell types are the spherical bushy cell (SB), globular bush cell (GB), stellite (St), multipolar cell (M), octopus cell (O), globular cell (G), and fusiform cell (F). The typical tuning curve, rate-level functions, period histogram, and PST histogram for auditory nerve fibers are shown at the bottom from left to right respectively. In other areas of the cochlear nucleus the tuning curves exhibit areas of excitation (*E*), inhibition (*I*), and responses that may be a little of both (?). All of the histogram types shown in Figure 16.5 are found in the PST histograms of cochlear nucleus fibers. Sometimes more than one type of PST histogram is found in one region, such as is shown on the left for the SB cells of the AVCN and on the right for the DCN. Additional circuitry is shown as the dotted lines for the DCN. Figure courtesy of Bill Shofner, Parmly Hearing Institute, Loyola University of Chicago, adapted from the work of Young (1984).

So each cell type in the cochlear nucleus has a peculiar anatomy and circuitry that results in specific PST histogram, intensity response, and tuning curve. I want to highlight a few well-known cell types.

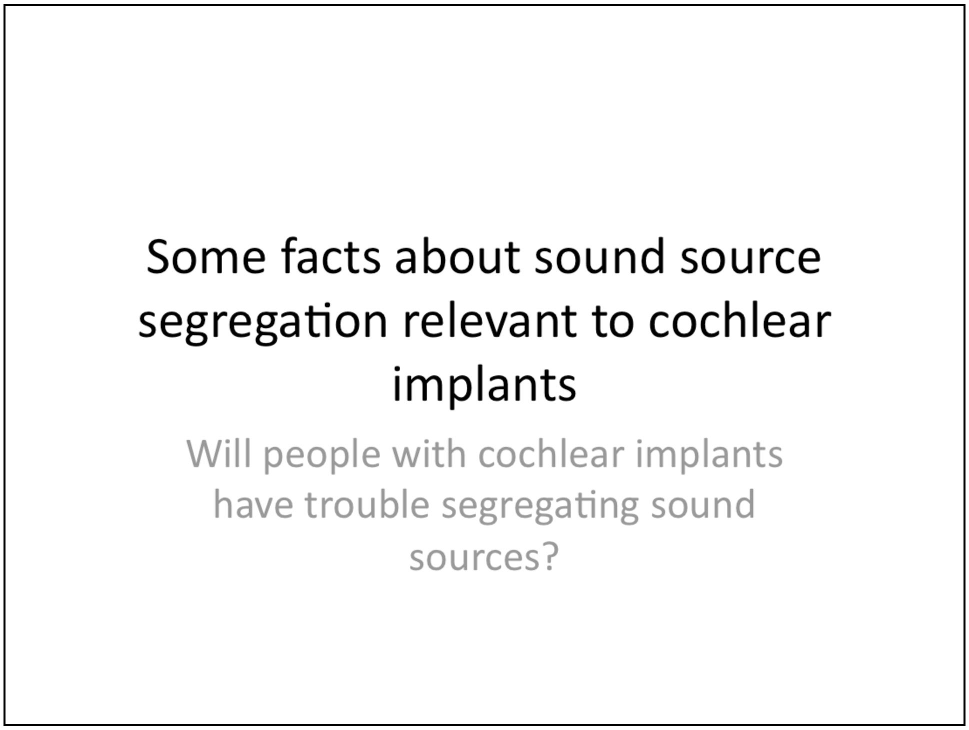
The octopus cells in the PVCN have a very interesting PST histogram (second graph from the top, below the tuning curve). The neuron produces a big onset response, and then it just stops responding. This is called an onset response. If you were trying to figure out if two sounds came on at the same time, this neuron isolates just the onsets.

The multipolar cells in the PVCN also have interesting PST histograms; the stellate cells in the AVCN have similar response properties. They have an onset response, and then they stop. Then they respond again, and stop, and again and stop. Their on-off behavior occurs at a regular rate that is different for different cells. The is called a chopper response. Chopper cells respond maximally to amplitude modulated sounds. So if you are looking for your temporal modulations, this would be these would be the neurons to go to.

Finally, look ath two neuron types in the DCN, giant cells and fusiform cells. They have different sorts of the PST histograms, but what is interesting about hem is their “tuning curves”, the top left graph. This doesn’t look like any tuning curve we’ve seen before. The neruon is inhibit by a braod range of frquencies and intenistites, and it is excited by only very specific frequency and intensity combinations– shown in the small dark pathces in the graph. This is a neuron that is looking for very specific spectral features.

Conclusions

- The auditory system analyzes sounds into their component frequencies and then segregates the components that belong together.
- Sound source segregation affects our ability to detect sound.
- Sound source segregation depends on the ability of the system to identify the spectrum, spectral shape (profile), amplitude envelope, pitch and location of sounds.
- The extraction of these features begins at the lowest level of the auditory pathways.


Some facts about sound source segregation relevant to cochlear implants
Will people with cochlear implants have trouble segregating sound sources?

It depends on whether they have the code for these cues

- Temporal separation
- Temporal onsets and offsets
- Temporal modulations
- Spectral profile
- Spectral separation
- Harmonicity
- Spatial separation